Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Systems Biology

Iterative Systems Biology for Medicine – Time for advancing from network signatures to mechanistic equations

David Gomez-Cabrero[a,b,c,d,e,f,1] and Jesper Tegnér[a,b,c,d,g,1]

Abstract

The rise and growth of Systems Biology following the sequencing of the human genome has been astounding. Early on, an iterative wet-dry methodology was formulated which turned out to be a successful approach in deciphering biological Complexity. Such type of analysis effectively identified and associated molecular network signatures operative in biological processes across different systems. Yet, it has proven difficult to distinguish between causes and consequences, thus making it challenging to attack medical questions where we require precise causative drug targets and disease mechanisms beyond a web of associated markers. Here we review principal advances with regard to identification of structure, dynamics, control, and design of biological systems, following the structure in the visionary review from 2002 by Dr. Kitano. Yet, here we find that the underlying challenge of finding the governing mechanistic system equations enabling precision medicine remains open thus rendering clinical translation of Systems Biology arduous. However, stunning advances in raw computational power, generation of high-precision multi-faceted biological data, combined with powerful algorithms hold promise to set the stage for data-driven identification of equations implicating a fundamental understanding of living systems during health and disease.

Addresses

[a] Unit of Computational Medicine, Department of Medicine, Karolinska Institutet, Stockholm, 171 77, Sweden
[b] Center for Molecular Medicine, Karolinska Institutet, Stockholm, 171 77, Sweden
[c] Unit of Clinical Epidemiology, Department of Medicine, Karolinska University Hospital, Stockholm, L8, 17176, Sweden
[d] Science for Life Laboratory, Solna, 17121, Sweden
[e] Mucosal and Salivary Biology Division, King's College London Dental Institute, London, SE1 9RT, United Kingdom
[f] Translational Bioinformatics Unit, NavarraBiomed, Departamento de Salud-Universidad Pública de Navarra, Pamplona, 31008, Navarra, Spain
[g] Biological and Environmental Sciences and Engineering Division, & Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, 23955–6900, Saudi Arabia

Corresponding authors: Gomez-Cabrero, David, Mucosal and Salivary Biology Division, King's College London Dental Institute, London, SE1 9RT, United Kingdom. (david.gomez_cabrero@kcl.ac.uk); Tegnér, Jesper, Biological and Environmental Sciences and Engineering Division, & Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, 23955–6900, Saudi Arabia. (jesper.tegner@kaust.edu.sa)
[1] Authors contributed equally.

Current Opinion in Systems Biology 2017, **3**:111–118

This review comes from a themed issue on **Clinical and translational systems biology (2017)**

Edited by **Jesper Tegnér** and **David Gomez-Cabrero**

For a complete overview see the Issue and the Editorial

Available online 8 May 2017

http://dx.doi.org/10.1016/j.coisb.2017.05.001

Keywords

Systems medicine, Kitano, Mathematical modeling, Causality.

In our view, Systems Biology has now become an accepted paradigm in biological research [1]. This is in part reflected in the sheer number and quality of publications utilizing systems approaches acknowledging and embracing the complexity of biology [2–4]. Recently, the application of such frameworks to clinical challenges has led to the emergence of what could be referred to as systems or network medicine [1,5,6]. It is therefore timely to ask *to what extent real progress has been achieved* – and to critically assess the nature of conceptual and technical hurdles remaining in meeting the needs from a medical standpoint. Here, we use the structure of the very influential position paper (close to 4000 citations) by Kitano in 2002 [7] to assess achievements and challenges on the basis of the research agendas put forward. Next, on the basis this analysis we argue that despite conceptual and technical advances, there remains a fundamental gap between finding associated features (biomarkers) of a given process versus the more challenging task obtaining a causal (e.g. mechanistic) understanding of the process. This, in our view an ultimate gap, becomes even more glaring in a medical context, since there we would like to ask therapeutic questions such as what happens if we do X to a (human) system. At the end of the day, X is an intervention based on causal understanding in the sense that "*if X is executed*" then "*the relevant processes become properly modified*". We conclude this opinion paper with the sentiment that the time is ripe for bridging this gap and algorithmic tools in combination with richer data and more powerful computational platforms have the potential to operationally address the inherent challenges in wordings such as 'relevant' and 'properly' above.

Systems-based analysis a la Kitano

Since the sequencing of the human genome, there has been a shift in biomedical research from reductionism towards a holistic view in the sense of acknowledging the complexity and myriad of parallel and interconnected processes, including the multiple spatio-temporal scales involved in almost any biological phenomena. Interestingly, technological advances rather than theory itself have largely driven this shift of perspective. It has generated a multitude of novel methodologies (or creative applications of existing methodologies), many of them labeled under the fields of Systems Biology [7] or Systems Medicine [8,9]. While multiple complementary definitions of Systems Biology do exist [10,11], we frame our discussion using the landmark paper from Prof. Kitano in 2002 [7]. Prof. Kitano provided a comprehensive concept, and what could be referred to as a normative account, in turn translated into an operational *pipeline* defining Systems Biology as a methodology to understand biological systems. Specifically, an iterative standpoint was formulated such that a cycle of research combining dry-lab and wet-lab efforts would generate, validate or reject a hypothesis, and finally incorporated the outcomes of the analysis in the state-of-the-art amenable for a new iteration of the cycle. In this, Prof. Kitano emphasized four necessary vital avenues of investigations that jointly would admit system-level understanding: (1) *system structures* (for instance the network of interactions), (2) *system dynamics* (mathematical description and analysis), (3) *control method* (identification of the biological targets that can modulate or control the state of the cell) and a (4) *design method* (aiming to construct systems de novo to make use or to validate properties identified or hypothesis generated). Remarkably, in hindsight the 2002 Kitano's vision has turned out to be truly predictive in that we have witnessed remarkable progress in all those four areas, yet at different pace, and in part evolving in separate communities. For example, the emergence of the young dynamic filed of synthetic biology can be viewed as response to the need for design, which in turn can be traced back to Feynman's classic dictum on what you can't create you don't understand. At this juncture, we could conceptually ask whether these four areas are necessary, sufficient, or both to achieve systems understanding [12]. To shed light on this issue we will first briefly review progress in respective area, finding that the aforementioned gap between biomarkers and mechanisms cuts across all four areas (see Fig. 1).

Structure, dynamics, control, and design – progress and gaps

In engineering, or more specifically control theory, system identification is defined as a method for developing mathematical and computer-based models that represent the characteristics of that system from measurements of the system inputs and outputs [13]. Traditionally, linear systems have been in focus and the mathematical model captures the transfer function between input and output, thus not necessarily incorporating neither the underlying biophysical components nor the non-linear dynamics governing the interactions between the components over time. In contrast, in biology we aspire to identify not only the structure of cellular networks but also their dynamics, in order to achieve engineered control of the system [14,15]. This motivates the division of labor between finding the structure, dynamics, and control respectively as originally conceptualized by Kitano. The *identification of* ***System Structures*** can be attained by data-driven reverse-engineering approaches [16], either augmented by prior knowledge as a structural scaffold or by direct experimental analysis requiring structural learning directly from the data [17]. With the advent of high-throughput technologies – including both microarray and Next-Generation Sequencing technologies, reverse-engineering approaches have been a major research area in Systems Biology since the original 2002 publication. Pure data-driven reverse-engineering methods have as a rule only used time-series and/or perturbation experiments to uncover associations – not necessarily causal – between features e.g a transcription factor and the expression of the corresponding target genes [18]. Such relationships can readily be represented using different modeling formalisms, such as Mutual information [11], Boolean networks, Bayesian networks (BNs) [12], Petri nets [19], constraint-based models, differential equations [20], rule-based models [21], cellular automata or agent-based models [22], all being parts of a growing toolkit for data-driven reverse-engineering approaches. Yet, causal parameterizing remains challenging due to uncertainties in model structure and parameters [23]. A second line of reasoning is to define a prior network structure or scaffold through a literature review. Examples include modeling of atherosclerosis modeling [20], brain functioning [24], or immune system [25] to name a few examples among many. Alternative, the prior structural template can be collected from systematic experiments, as in the case of Protein–Protein interactions and the generation of the Proteome-Scale Map of the Human Interactome Network [17]. From the three approaches, experimental and data-driven approaches are in our view to become even more prevalent due to the exponential growth of data in public repositories [26–28] and the decrease cost in sequencing [29]. The knowledge-based approach appears to be at a turning point in the sense that "classical" text-mining methodologies [30,31] have not, in our view, provided a significant edge when compared with other approaches, whereas recent advances using DeepLearning [32] methodologies hold promise to disrupt current state-of-the-art in text mining similarly to recent achievements in genomic analysis [33,34]. In summary, these advances in network biology have enriched the notion of biomarker from a single or very few features to include a larger set of (inter-connected) features (i.e. a network signature) associated with